INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Domain-initial strengthening in Turkish: Acoustic cues to prosodic hierarchy in stop consonants

Kubra Bodur, Sweeney Branje, Morgane Peirolo, Ingrid Tiscareno, James S. German

Aix-Marseille Université, LPL, CNRS, Aix-en-Provence, France
kubra.bodur@etu.univ-amu.fr, james.german@lpl-aix.fr

Abstract

Studies have shown that cross-linguistically, consonants at the left edge of higher-level prosodic boundaries tend to be more forcefully articulated than those at lower-level boundaries, a phenomenon known as *domain-initial strengthening*. This study tests whether similar effects occur in Turkish, using the Autosegmental-Metrical model proposed by Ipek & Jun [1, 2] as the basis for assessing boundary strength. Productions of /t/ and /d/ were elicited in four domain-initial prosodic positions corresponding to progressively higher-level boundaries: syllable, word, intermediate phrase, and Intonational Phrase. A fifth position, nuclear word, was included in order to better situate it within the prosodic hierarchy. Acoustic correlates of articulatory strength were measured, including closure duration for /d/ and /t/, as well as voice onset time and burst energy for /t/. Our results show that closure duration increases cumulatively from syllable to intermediate phrase, while voice onset time and burst energy are not influenced by boundary strength. These findings provide corroborating evidence for Ipek & Jun's model, particularly for the distinction between word and intermediate phrase boundaries. Additionally, articulatory strength at the left edge of the nuclear word patterned closely with word-initial position, supporting the view that the nuclear word is not associated with a distinct phrasing domain.

Index Terms: Turkish, domain-initial strengthening, intonational phonology, prosody

1. Introduction

Standard Turkish (ST) is generally characterized as a stress accent language where each word has a lexically determined stressed syllable capable of attracting a pitch accent [3, 4, 5]. This syllable is most often final in the word, though certain morphological patterns shift the stress to earlier positions. As discussed in [1, 2], a typical declarative sentence in ST is realized with a series of f0 rises, which the authors account for through a low tone (L) which marks the left edge of each prosodic word, and a high pitch accent (H*) which associates to the lexically stressed syllable of each word. Additionally, a rising tone (LH-) marks the left edge of each intermediate phrase (ip). Since stress is typically word-final, LH- is most often coincident with H*, resulting in just a single rise to the end of the ip-final word. When stress is non-final however, H* is realized separately from LH- resulting in a rising-falling-rising pattern over the ip-final word. For utterances with canonical SOV word order in neutral contexts, the pre-verbal object has the status of "nuclear word" (Wn) and is marked with a downstepped pitch accent (!H*) followed by postnuclear compression (deaccenting). Additionally, as shown by [1, 2], this position is marked by a rising f0 (LHn) aligned to the right edge of the word immediately preceding it. As with LH-, Hn tends to be coincident with H* unless stress on the prenuclear word is non-final.

The Autosegmental-Metrical model of [1, 2] implies a prosodic hierarchy progressing from syllable to (prosodic) word to ip to intonational phrase (IP). (See Figure 1). As this model is based on evidence from f0, it is useful to seek independent sources of evidence from other potential markers of prosodic boundary strength which corroborate it. Studies have shown that in many languages (e.g. Korean, French, Taiwanese, English), onset consonants bear articulatory and acoustic markers of the relative strength of the prosodic boundary at whose left edge they occur. These markers include greater lingo-palatal contact area, longer articulatory seal duration, longer acoustic closure duration, and longer voice onset time (VOT) [6]. [7] additionally found that in Korean, higher-level boundaries are weakly associated with *lower* acoustic energy in the release burst, though only for certain consonants. As suggested by [8] and discussed in [7], this could be an artifact of increased linguo-palatal contact in higher domains, which results in a longer release duration and therefore a lower peak in burst energy.

Crucially, the relationship between prosodic boundary strength and the various phonetic measures mentioned above is specific to both languages and consonant categories within those languages. Not only do different languages have different prosodic hierarchies involving different structures, but languages also differ in terms of which phonetic parameters are recruited for marking those structures. VOT, for example, is sensitive to prosodic position in Korean, but not in French or Taiwanese [6]. Within Korean, the VOT of /t^h/ but not /t*/ is positively correlated with boundary strength, while only the latter phoneme shows a negative correlation between burst energy and boundary strength. Furthermore, /t*/ is actually *longer* in word-medial contexts as compared to word-initial contexts, thereby going against the more general trend of cumulative strengthening [7]. In short, it is difficult to predict for a given language which phonetic measures are associated with prosodic boundary strength, and to the extent that they are, which levels of the prosodic hierarchy are concerned.

Nevertheless, evidence for domain-initial strengthening, when it occurs, is a useful tool for validating models of prosodic organization based on intonational patterns. In this study, we therefore seek evidence for an association between articulatory strength and the prosodic boundary levels proposed in [1, 2]'s model of ST intonation. We assume that, to the extent that domain-initial strengthening exists in ST, it is cumulative, i.e., that consonants at the left edge of increasingly higher prosodic domains will have increasingly stronger articulatory properties with respect to one or more measures. Here, we focus on acoustic correlates of articulatory

http://dx.doi.org/10.21437/Interspeech.2021-2230

strengthening for two stop consonants, including closure duration for /d/ and /t/, VOT for /t/, and burst energy for /t/ as measured by the root-mean-square (RMS) of the acoustic signal at the closure release. Based on findings from other languages, our specific hypothesis is therefore that higher positions in the prosodic hierarchy proposed by [1, 2] will be associated with longer durations, longer VOTs, and reduced burst energy as compared to lower positions. The principal prosodic boundary levels we test include the syllable, the word, the ip, and the IP (in this case, also an utterance). Since one of the main contributions of [1, 2]'s model is the need for an ip-level domain, domain-initial effects that differentiate the word and ip levels are of key interest to our study.

Additionally, [1, 2] refer to LHn as a "boundary tone", though they stop short of associating it with a specific domain within the hierarchy. In order to assess whether the juncture between the pre-nuclear word and the nuclear word (henceforth, the Wn boundary) has its own status within the prosodic hierarchy of ST, we also include that position in our study. Since the Wn boundary is minimally also a word boundary, we do not expect it to be articulated with less strength than word boundaries. It is possible, however, that it is phonologically equivalent to an ip boundary, in which case it should pattern with ip, or that it has a unique status, in which case it may pattern between the word and the ip or possibly even be articulated with greater strength than ip.

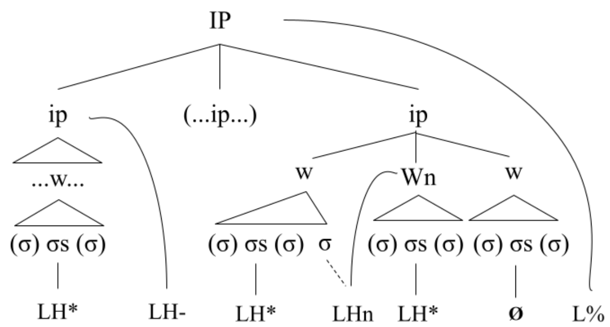

Figure 1: *Schematized prosodic hierarchy of Turkish following Ipek & Jun's model. 'σs' indicates lexically stressed syllables.*

2. Methods

2.1. Participants

Eight female native speakers of ST (mean age = 39.8; SD = 8.9) participated in the study. All participants were living in Marseille, France at the time of the study and had resided in France for less than 5 years. While they were not all originally from Istanbul, a native speaker who is one of the authors verified that they all spoke a variety consistent with ST.

2.2. Materials

Six sets of ten target sentences were created. Each set included five words targeting /t/ and five words targeting /d/ in each of the prosodic positions tested. This was achieved by constructing sentences which favored specific intonational patterns that located the target consonants in the desired positions. Target words within each set were matched as closely as possible given the constraints imposed by morphosyntax. In most cases, target words within a set and for a given consonant were morphologically related for the word, Wn, ip, and IP (or utterance) conditions. Words in the syllable condition were matched to the other conditions for the vowel following the target consonant. An example of the /d/ portion of one set is shown in Table 1.

Table 1. *Target sentences for the /d/ portion of one item set.*

Prosodic boundary	Experimental sentence
s (syllable)	Kıraathanenin mü**da**vimleri çoğunlukla *(Coffee house's) (regulars) (mostly)* emeklilerdi. *(were retirees).*
w (word)	Herkes bu **davada** kimin *(Everybody) (this) (in the case) (who)* haklı olduğunu çözmeye çalışıyordu. *(was right) (to figure out) (was trying).*
Wn (nuclear word)	Genç avukat hayatında ilk kez bir *(Young) (lawyer) (in his life) (the first) (time) (a)* **davaya** katıldı. *(to lawsuit) (attended).*
ip (intermediate phrase)	Savcı **davayı** haftaya *(Prosecutor) (the lawsuit) (next week)* bir meslektaşına devredecek. *(to a colleague) (will hand over).*
IP (intonational phrase)	**Dava** sonuçlanmadan size *(The lawsuit)(before concluded)(to you)* bir şey söylemem doğru olmaz. *(something) (for me to tell) (wouldn't be right).*

The 60 target sentences were first assigned to six counterbalanced blocks of ten items including five /t/ items and five /d/ items. The six blocks were then pseudorandomized in four different orders for a total of four experimental lists. Furthermore, the order of items within each block was randomized separately for each list.

2.3. Procedures

Two participants were recorded in a sound attenuated booth at the Laboratoire Parole et Langage in Aix-en-Provence. Due to restrictions related to Covid-19 countermeasures, the other six were recorded in their homes in a quiet room accompanied by one of the experimenters. All recordings were conducted using the same microphone (Fifine K668) captured with Audacity. Participants were seated with a laptop in front of them, and target sentences were presented one at a time on the screen. Due to protocol requirements related to Covid-19, head-mounted microphones could not be used. However, measures were taken to ensure that microphones were placed at a similar and stable distance (approx. 60 cm) for all participants. Participants were asked to first read each sentence silently to understand its meaning, and then to read it aloud as naturally and conversationally as possible. They were permitted to repeat a sentence in case of hesitations or speaking errors. Before beginning the experiment and in order to verify participants' understanding of the task, they were asked to produce three practice items which did not include the phonemes /d/ or /t/.

2.4. Data analysis

Prior to acoustic analysis, all recorded trials were analyzed and classified for the type of prosodic boundary at the target location using visual and auditory inspection in Praat [9]. Each trial was analyzed by two different authors and the resulting classifications were compared. In case of disagreement, the authors collectively discussed the pattern to arrive at a final classification. When a consensus could not be reached, the trial was discarded. Additionally, following [7], productions where a pause occurred before the target word were excluded, due to the difficulty of identifying the onset of a closure following a silence. For the same reason, duration was not measured for the IP boundary condition.

Each trial was annotated in Praat by one of the authors for the onset and the offset of the closure of the target consonant, as well as for the location of VOT for /t/. This annotation closely followed the guidelines outlined in [10]. VOT was not analyzed for /d/ due to the fact that virtually all trials were fully voiced, resulting in only a small number of tokens that would yield meaningful measurements. Duration of the closure, VOT, and burst energy were then extracted automatically using a custom Praat script. Following [7], burst energy was measured as total RMS within a 10 ms window centered on the release of the closure. RMS for /d/ was not measured due to continuous voicing the release. Since calibrated head-mounted microphones could not be used, RMS values were z-transformed by speaker.

3. Results

A total of 480 target productions were collected. 34 items (/d/: 20, /t/: 14) were excluded due to the presence of a pause preceding the target word. An additional three items (/d/: 1, /t/: 2) were excluded because the prosodic boundary at the target was not the expected one. A total of 443 items were therefore used for analysis.

3.1. Closure duration

Figure 2 shows boxplots of closure duration for /t/ for four prosodic positions. Overall, duration appears to increase from syllable to word to ip, while Wn appears to pattern closely with word.

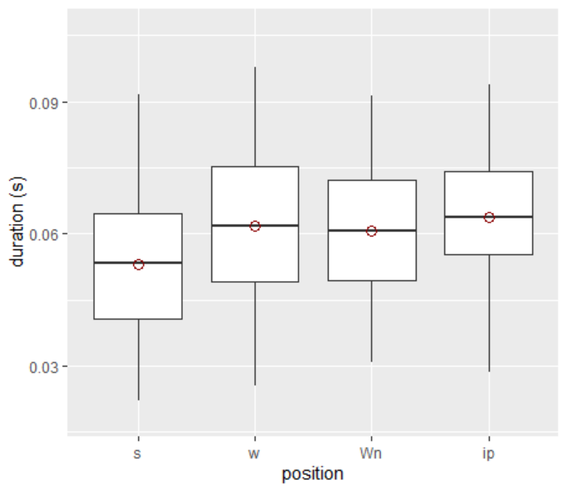

Figure 2: *Boxplot of /t/ closure duration for four positions.*

To assess these differences, a linear mixed regression model was fit to the data, including position as a fixed factor, and participants and items as random effects. A log-likelihood ratio test comparing the full model to a null model revealed a significant main effect of position ($\chi2(3) = 19.0$, $p < 0.001$). To further explore the source of this effect, a set of post hoc pairwise comparisons with Tukey correction were carried out. These revealed significant differences between syllable and each of the other levels (word: $z = 2.91$, $p < 0.05$; Wn: $z = 2.68$, $p < 0.05$; ip: $z = 4.32$, $p < 0.001$) with no further significant comparisons.

Figure 3 shows a similar pattern for /d/. The same statistical procedure used above again revealed a significant main effect of position on duration ($\chi2(3)=35.6$, $p < 0.001$). As before, pairwise post hoc comparisons revealed significant differences between syllable and each of the other levels (word: $z = 2.74$, $p < 0.05$; Wn: $z = 2.65$, $p < 0.05$; ip: $z = 6.24$, $p < 0.001$). Crucially, however, significant differences were also found between word and ip ($z = 3.38$, $p < 0.01$), and between Wn and ip ($z = 3.56$, $p < 0.01$), while word and Wn did not differ significantly ($z = 0.13$, $p = 0.99$). This result not only provides support for the relevance of the ip domain in ST, but also suggests that Wn is equivalent to a word boundary.

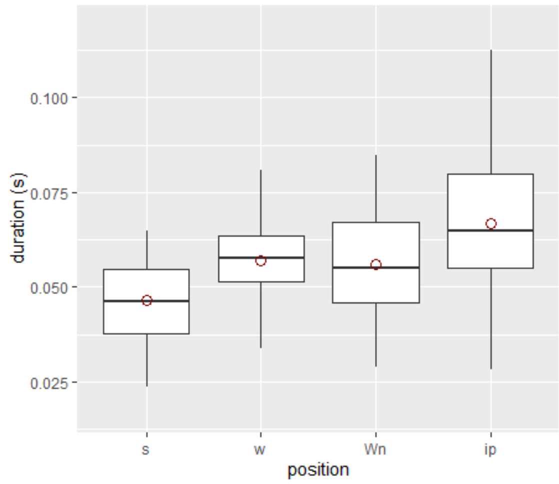

Figure 3: *Boxplot of /d/ closure duration for four positions.*

3.2. Voice Onset Time

Figure 4 shows VOT for /t/ for five prosodic positions. Descriptively, these results do not show any particularly salient trend. Indeed, a linear mixed effects regression analysis (as above) revealed no significant effect of position for VOT ($\chi2(4)=2.96$, $p = 0.56$).

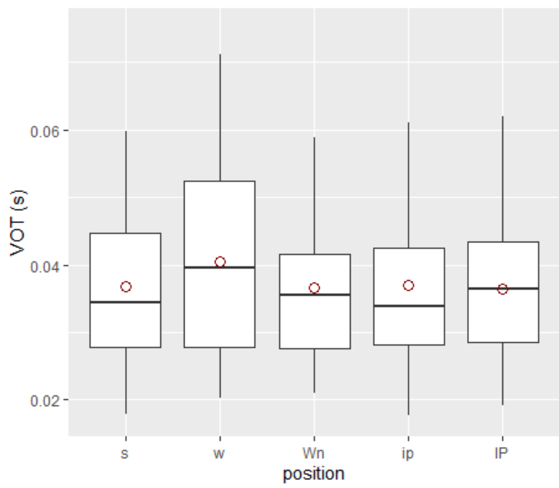

Figure 4: *Boxplot of VOT of /t/ for five positions*

3.3. Burst energy

Figure 5 shows burst energy in RMS by prosodic position for /t/. Descriptively, the only salient feature is that the IP level appears to be lower than the others. A linear mixed model including item as the only random effect (since z-transformation was carried out over subjects) with log-likelihood ratio comparison to the null model revealed that the effect of position is indeed significant ($\chi 2(4) = 14.4$, $p < 0.01$). Post hoc pairwise comparisons revealed that this effect is likely due to a lower mean RMS for IP, since that position was significantly different from syllable ($z = -3.31$, $p < 0.01$), from word ($z = -2.85$, $p < 0.05$), from Wn ($z = -2.99$, $p < 0.05$), and marginally different from ip ($z = -2.48$, $p = 0.095$), with no other significant comparisons. While the trend is not cumulative, this result is consistent with our predictions and the findings of [7] for Korean.

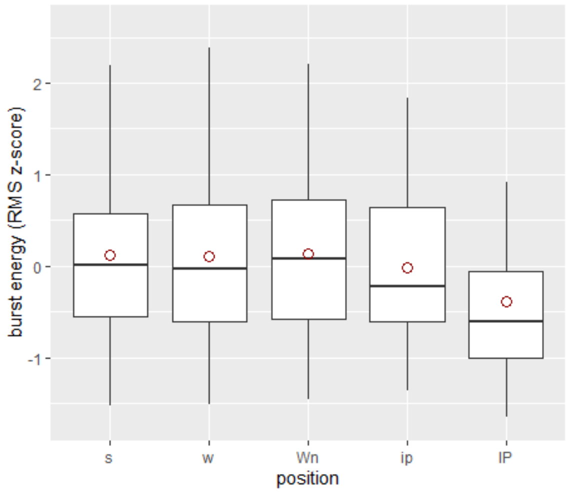

Figure 5: *Boxplot of burst energy of /t/ for five positions.*

4. Discussion

In order to assess and/or corroborate the proposal of [1,2] regarding prosodic hierarchy levels in ST, we tested the role of prosodic boundary strength on the acoustic realization of the consonants /t/ and /d/ in domain-initial positions. In this regard, evidence for *cumulative* strengthening was found only with respect to closure duration, and only for /d/. Nevertheless, the overall pattern including a three-way distinction between syllable, word, and ip provides important corroborating evidence for the phonological relevance of these levels in the prosodic system of Turkish. Additionally, our finding that Wn patterns very closely with the word-level suggests that the left edge of the nuclear word, while important for the assignment of postlexical tones, does not appear to project an ip-level boundary, and likely does not represent a prosodic domain distinct from the word-level. While closure duration for /t/ showed a less robust pattern, the overall trend is highly similar to that for /d/. It is possible that the lack of significance across levels is due to the relatively small number of participants and/or the non-ideal conditions of data collection which were necessitated by restrictions related to Covid-19. We therefore suspect that a larger sample of participants tested under laboratory conditions would elicit more robust effects overall for duration.

Additional support for domain-initial strengthening was found in the reduced burst energy associated with the IP-level for /t/. While closure duration was not measured in that position, the mechanism proposed by [8] suggests that the lower burst energy is the result of stronger lingo-palatal contact and longer closure durations as compared to the other positions, which is consistent with the cumulative nature of duration in our results for /d/. We found no particularly revealing pattern in the results for voice onset time. However, this is not incompatible with the fact that we found positive evidence for progressive strengthening in the closure duration domain, since each language recruits different phonetic parameters in different ways. Our results simply mean that duration is likely an important cue for prosodic position in ST while VOT is not.

During the analysis of our data, we encountered a phenomenon that we did not expect, and which could have potential implications for domain-initial strengthening in Turkish. Specifically, we observed that /d/ was often produced with a long, apparently aspirated, interval following the release of the closure. As shown in Figure 6, in such cases the closure itself was generally either fully or partially voiced, and voicing bars were clearly present in the spectrogram during this post-release interval in conjunction with a large amount of high frequency energy. This interval ends with the onset of fully modal voicing, which includes a sudden increase in periodic energy and a sudden disappearance of the higher frequency energy (aspiration). In short, it appears that at least some ST speakers produce /d/ as a murmured stop, and that this voice quality persists for a certain, and possibly systematic, interval after the release of the closure, much like the delay in the onset of voicing associated with typical voiceless aspirated stops. To our knowledge, no case of a murmured stop has been reported in literature regarding Turkish phonology. In the present study, we did not measure this interval. However, just as /d/ and /t/ differ in how their duration is related to prosodic position, it is possible that the length of the post-release murmured interval for /d/ may be associated with prosodic position even though the VOT for /t/ is not.

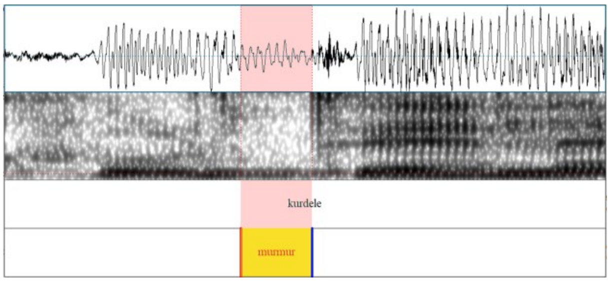

Figure 6: *Example of "murmured" /d/ at a syllable boundary (*kur<u>d</u>ele *'ribbon').*

5. Acknowledgements

We would like to thank the students of HSLCU10 - *Phonologie de laboratoire* for their contribution during different stages of this study : Yassamine Benmehidi, Auriane Boudin, Agathe Codron, Charlie Hallart, Xizi Liu, Aya Remmache, Jack Wood and Shiyang Zhu. This research was partially supported by grants ANR-16-CONV-0002 (ILCB) and the Excellence Initiative of Aix-Marseille University (A*MIDEX).

6. References

[1] Ipek, C., & Jun, S.-A. (2013). Towards a model of intonational phonology of Turkish : Neutral intonation. The Journal of the Acoustical Society of America, 133(5), 3573.

[2] Ipek, C., Jun, S. (2014) Distinguishing Phrase-Final and Phrase-Medial High Tone on Finally Stressed Words in Turkish. Proc. 7th International Conference on Speech Prosody 2014, 393-397

[3] Lees, R. (1961). *The phonology of Modern Standard Turkish*. Bloomington: Indiana University.

[4] Lewis, G. (1967). *Turkish Grammar*. Oxford: OUP.

[5] Sezer, E. (1981). On non-final stress in Turkish. *Journal of Turkish Studies*, 5, 61-69.

[6] Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2004). Domain-initial articulatory strengthening in four languages. Phonetic interpretation: Papers in laboratory phonology VI, 143-161.

[7] Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. Journal of Phonetics, 29(2), 155 - 190.

[8] Stevens, K. N., Keyser, S. J. & Kawasaki, H. (1986). Toward a phonetic and phonological theory of redundant features. In Perkell, J. S. & Klatt, D. H. (eds), *Invariance and variability in speech processes.* NJ: Lawrence Earlbaum Associates.

[9] Boersma, Paul & Weenink, David (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.41, retrieved 1 April 2021 from http://www.praat.org/

[10] Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. Methods in empirical prosody research, 3, 1-28.